

UNITED STATES PATENT APPLICATION

**SPEECH TO TEXT CAPTIONING DIGITAL CAMERA**

**INVENTOR**

**John W. Sherry**

66400T" 0044F460

Schwegman, Lundberg, Woessner & Kluth, P.A.  
1600 TCF Tower  
121 South Eighth Street  
Minneapolis, MN 55402  
ATTORNEY DOCKET SLWK 884.166US1  
Client Reference P7377

## Speech to Text Captioning Digital Camera

5  
A  
ind  
B1

### Technical Field of the Invention

The present invention relates generally to digital cameras and in particular the present invention relates to captioning speech in a digital camera.

### Background of the Invention *information*

10 The field of photography has experienced significant advancements with the development of digital cameras. These cameras capture images and store the images as a digital file. Some cameras include features to allow for text to be superimposed on images. For example, a date or caption such as "1999 vacation" can be entered and superimposed on pictures. This text is typically used for several pictures and changes  
15 are relatively difficult.

In a similar manner, some cameras include a feature to capture speech. The speech is stored as a separate audio file from the image file. The two files are usually named in a similar manner (different extension), such as 0021.jpg and 0021.mpa for respective image and speech files for picture 0021. Playback of the audio file is  
20 awkward on a device other than the camera.

For the reasons stated above, and for other reasons stated below which will become apparent to those skilled in the art upon reading and understanding the present specification, there is a need in the art for a camera which allows for custom picture annotation using an audio input.

25

### Summary of the Invention

In one embodiment, a camera comprises a photo-sensitive array for capturing an image, a microphone, a memory, and a processor coupled to the photo-sensitive array,  
30 microphone and memory. The processor converts audio input provided by the microphone into text and stores the text in the memory.

662007 "004466

In another embodiment, a camera comprises a photo-sensitive array for capturing an image, a microphone, a memory, and a processor coupled to the photo-sensitive array, microphone and memory. The processor converts captured audio input provided by the microphone into a digital text file and converts the captured image into a digital image file. The processor also stores the digital image file and the digital text file as a single composite digital data file in the memory.

A method of operating a camera is provided which comprises activating a shutter of the camera to capture a light image, and converting the light image to digital image data. The method also includes activating an audio input, capturing audio input, and converting the audio input into text data.

#### Brief Description of the Drawings

Figure 1 is a block diagram of a camera according to one embodiment of the present invention;

Figure 2 is an example photograph including a text message;

Figure 3 is a flow chart of an example operation of the camera of Figure 1; and

Figure 4 illustrates an alternate embodiment of a camera system of the present invention.

#### Detailed Description of the Invention

In the following detailed description of the preferred embodiments, reference is made to the accompanying drawings which form a part hereof, and in which is shown by way of illustration specific preferred embodiments in which the inventions may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized and that logical, mechanical and electrical changes may be made without departing from the spirit and scope of the present invention. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the claims.

Figure 1 illustrates a digital camera 100 according to one embodiment of the present invention. The camera includes an input lens 102 (including a shutter) for directing light to a light sensitive array 104. In operation the shutter is activated to direct a light image onto the array. This array can be any photosensitive array, such as a charge coupled device (CCD). CCDs have dominated vision applications because of their superior dynamic range, low fixed-pattern noise and high sensitivity to light. The array is coupled to a processor 106. The processor can transfer the image captured in the array to memory 110. Numerous images stored in the memory can be output through output circuitry 112. A microphone 108 is coupled to the processor for capturing speech, as explained below. The processor can process the speech and then route the speech to the memory for storage. The term processor as used herein encompasses circuitry required to convert the light image to digital data. Further, the processor can include any control circuitry needed to operate the camera, including output operations.

In operation, a user of the digital camera can activate a speech capturing operation using control 114. In one embodiment, control 114 comprises a switch which can be activated when speech is desired. Other types of control circuitry are contemplated, including speech recognition using the microphone and processor. In this embodiment, a pre-determined audio command is used to "wake-up" the camera speech capture mode. When the processor is in the speech capture mode, output speech signals provided by the microphone are converted to text using the processor. As such, the processor operates a speech translation operation. This operation is performed by executing a set of instructions which instruct the processor to convert the received speech. The text is then used to add captions to the photograph which was taken most recently.

*ind AB* Referring to Figure 2, a sample photograph is provided. The photograph includes a text message which describes the photograph as "sunset in the mountains during vacation of 1999". This text message is intended to illustrate the flexibility of the present invention to customize text on each photograph. The processor 106 can store the text message as either a separate text file (text data), or combine the

AB  
photograph data (digital image) and the text data as one digital file. As separate files, the speech can be stored as an audio file, or converted to text. By combining the files, unwanted separation of the photo and text can be avoided. Further, the text can be separately stored in parallel to the photo/text file to provide a level of redundancy.

- 5 Figure 3 provides an illustration of the operation of the digital camera. A user takes a photograph at 200 by activating the shutter. The photograph image is stored at 202. The user initiates audio speech input at 204, and the speech is converted to text at 206. Alternatively, a compressed audio file is stored of the speech at 208. The text is combined with the image at 210 to provide a common file which is stored in memory.
- 10 Alternately, the text is stored as a separate file at 212.

- Figure 4 illustrates a camera system of the present invention. The system includes a digital camera 300 and an external processor 310. The camera is substantially similar to the camera of Figure 1 except the speech to text processor is external to the camera. The camera includes a microphone to capture speech. The
- 15 speech and digital image are communicated to the external processor as separate files. The processor converts the speech to text, as explained above, and combines the image and text into a single data file. This common file can be stored by the processor. This system can provide the benefits explained above without the physical limitations of a portable camera. As such, the system is better suited to a studio setting. Again, this
- 20 process allows for the combination of text and image into a common file.

A ~~Conclusion~~

- A digital camera has been described which includes a processor and a microphone. The camera captures audio speech and converts the speech to text. The
- 25 text can be combined with a captured image to provide a composite file. The processor has been described as executing an instruction set which preforms the audio to text conversion.

- 30 Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement which is calculated to achieve the same purpose may be substituted for the specific embodiment

*all  
ended*

shown. This application is intended to cover any adaptations or variations of the present invention. Therefore, it is manifestly intended that this invention be limited only by the claims and the equivalents thereof.

662007-00447460